

The Processing of Temporal Pitch and Melody Information in Auditory Cortex

Roy D. Patterson,^{1,5} Stefan Uppenkamp,¹
Ingrid S. Johnsrude,² and Timothy D. Griffiths^{1,3,4}

¹Centre for the Neural Basis of Hearing
Physiology Department
University of Cambridge
Downing Street
Cambridge CB2 3EG

²MRC Cognition and Brain Sciences Unit
15 Chaucer Road
Cambridge CB2 2EF

³Wellcome Department of Imaging Neuroscience
Institute of Neurology
12 Queen Square
London WC1N 3BG

⁴Auditory Group
Newcastle University Medical School
Newcastle upon Tyne NE2 4HH
United Kingdom

Summary

An fMRI experiment was performed to identify the main stages of melody processing in the auditory pathway. Spectrally matched sounds that produce no pitch, fixed pitch, or melody were all found to activate Heschl's gyrus (HG) and planum temporale (PT). Within this region, sounds with pitch produced more activation than those without pitch only in the lateral half of HG. When the pitch was varied to produce a melody, there was activation in regions beyond HG and PT, specifically in the superior temporal gyrus (STG) and planum polare (PP). The results support the view that there is hierarchy of pitch processing in which the center of activity moves anterolaterally away from primary auditory cortex as the processing of melodic sounds proceeds.

Introduction

This paper is concerned with three auditory processes involved in the perception of melody, and how these processes are organized in the ascending auditory pathway. A melody in this case is simply a sequence of notes like that produced when someone picks out a tune on the piano with one finger. From the auditory perspective, perception of a melody involves (1) detecting that segments of an extended sound contain temporal regularity, (2) determining the pitch of each of these regular segments, and (3) determining how the pitch changes from note to note over the course of the sound. Physiological studies (Palmer and Winter, 1992) and functional neuroimaging (Hall et al., 2002; Griffiths et al., 2001; Wessinger et al., 2001) suggest that the processing of temporal regularity begins in the brainstem and that pitch extraction is completed in HG, the site of primary auditory cortex (PAC) (Rademacher et al., 1993, 2001; Morosan

et al., 2001; Rivier and Clarke, 1997; Hackett et al., 2001). This is consistent with the hierarchy of processing proposed for auditory cortex on the basis of recent anatomical studies in the macaque (Hackett et al., 1998, 2001; Kaas and Hackett, 2000; Rauschecker and Tian, 2000). Higher-level processes like pitch tracking and melody extraction are thought to be performed in more distributed regions beyond PAC, and the processing becomes asymmetric with more activity in the right hemisphere (see Zatorre et al., 2002, for a review). The current paper presents cortical data from a functional magnetic resonance imaging (fMRI) study designed to increase the sensitivity of auditory imaging and enable us to locate the neural centers involved in pitch and melody perception with much greater precision.

Studies of pitch processing often employ sinusoids that activate focal regions on the basilar membrane; these studies show that the tonotopic organization observed in the cochlea is preserved in all of the nuclei of the auditory pathway up to PAC (for a review see Ehret and Romand, 1997). It is also possible to produce a tone with a strong pitch by regularizing the time intervals in a broadband noise so that one time interval occurs more often than any of the others (see Figure 1). As the degree of regularity increases, the hiss of the noise dies away, and the pitch of the tonal component increases to the point where it dominates the perception. These regular-interval (RI) sounds (Yost, 1998) are like noise insofar as they produce essentially uniform excitation along the basilar membrane and, thus, uniform activity across the tonotopic dimension of neural activity in the auditory pathway (compare Figures 1B and 1G). The fact that they produce a strong pitch without producing a set of harmonically related peaks in the internal spectrum shows that pitch can be coded temporally as well as tonotopically in the auditory system. Figure 1 illustrates how the auditory system could extract the pitch information from RI sounds. A brief comparison of spectral and temporal models of pitch is presented in Griffiths et al., 1998.

RI sounds are useful in imaging because they enable us to generate sets of spectrally matched stimuli that enhance the sensitivity of perceptual contrasts in functional imaging. Their value was initially demonstrated by Griffiths et al. (1998), who used positron emission tomography (PET) to show that activation in HG increases with the temporal regularity of RI sounds and that, when the pitch changes over time, there is additional activation in STG and PP. The power was limited, however, by constraints on radiation dose, and the spatial resolution was poor compared to that of fMRI. For these reasons, the results were restricted to group data, and they are ambiguous with regard to the degree of asymmetry at different stages. Subsequently, Griffiths et al. (2001) showed that the combination of RI sounds and fMRI was sufficiently sensitive to image all of the subcortical nuclei of the auditory pathway simultaneously, provided the technique included cardiac gating (Guimares et al., 1998) and many replications of each stimulus condition. A contrast between the activation

⁵Correspondence: roy.patterson@mrc-cbu.cam.ac.uk

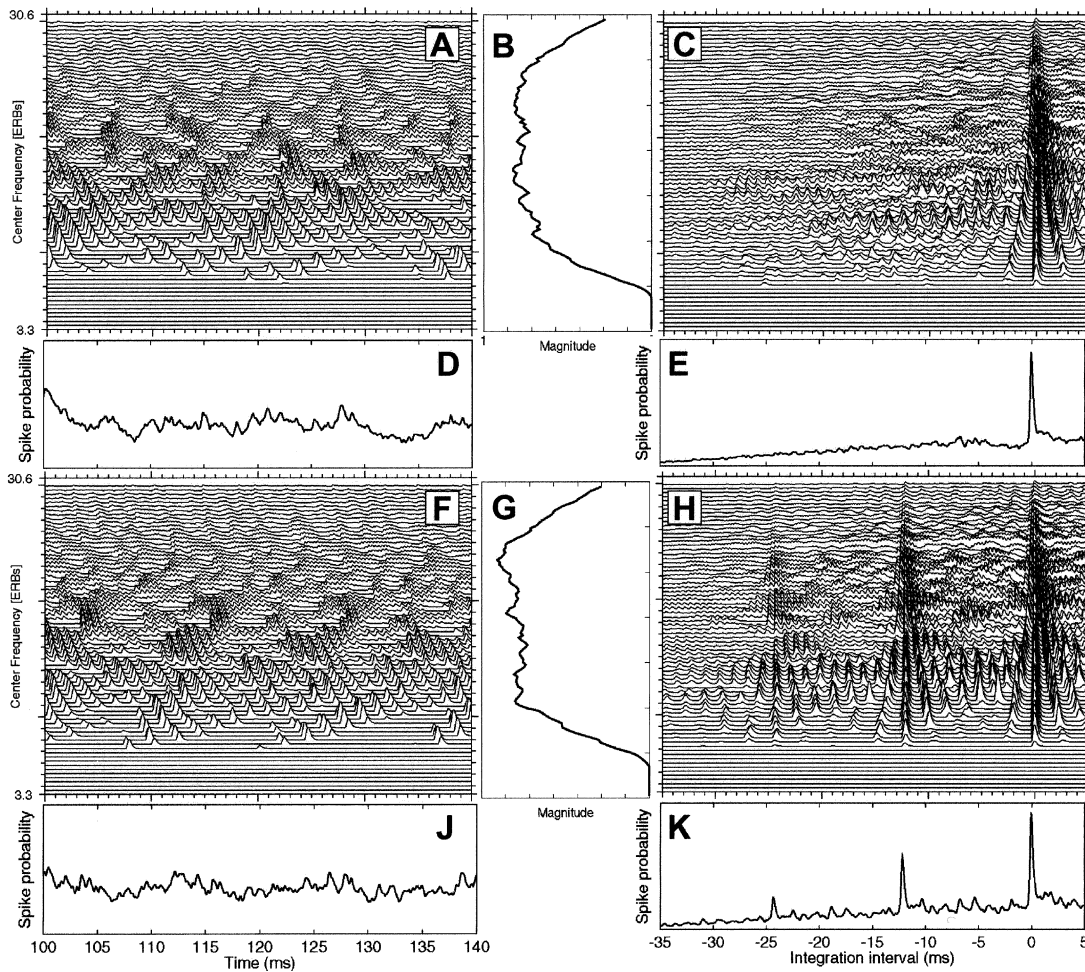


Figure 1. Simulated Neural Responses to a Random Noise and a Regular-Interval (RI) Sound with a Pitch of 83.3 Hz

The RI sound was constructed by: (1) delaying a copy of a random noise by 12 ms ($1/83.3$ Hz), (2) adding it back to the original noise, and (3) repeating the process 16 times (Yost, 1996; see Experimental Procedures for details). The ordinate in (A)–(C) and (F)–(H) is the tonotopic dimension of hearing, that is, place along the basilar membrane. (A) and (F) show the detailed temporal structure of the neural response at the output of the cochlea (Patterson et al., 1995) for the noise and RI sound, respectively; the activity is similar for the two sounds. (B) and (G) show that the average neural activity over time (Patterson, 1994) is very similar for the two sounds; there are no harmonically related peaks in (G) to identify the pitch. (D) and (J) show that the average activity over channels is also very similar; there are no regularly repeating features to identify the pitch of the RI sound. The temporal regularity that distinguishes the RI sound is in the time-interval information of the neural patterns (A and F). When the time-intervals between peaks are calculated in the individual channels of (F) and summarized in the corresponding time-interval histograms of (H), a concentration of activity appears at the RI delay (12 ms) (Patterson et al., 1996). In contrast, the time-intervals in the neural pattern of noise (A) are randomly distributed in the time-interval histogram (C) (the concentration at 0 ms simply indicates the presence of activity in the channel). The position and height of peaks in the average interval histogram (K) are used to evaluate quantitative models of the pitch of RI sounds (e.g., Pressnitzer et al., 2001; Krumbholz et al., 2000).

produced by RI sounds with fixed pitch and spectrally matched noise revealed that temporal pitch processing begins in subcortical structures. At the same time, a contrast between sounds with varying pitch and fixed pitch did not reveal an increase in activation in this region. The fact that pitch processing begins in the brainstem but is not completed there was interpreted as further evidence for the hypothesis that there is a neural hierarchy of melody processing in the auditory pathway (Griffiths et al., 1998). In this paper, we present the cortical data from the fMRI experiment. The exceptional sensitivity of the study enables us to track the hierarchy of melody processing in auditory cortex across HG and out into PP and STG, and determine

where the asymmetries reported by Zatorre et al. (2002) first emerge.

Studies of cytoarchitecture have shown that a reliable landmark for primary auditory cortex is the anteriormost, transverse temporal gyrus (of Heschl) (Rademacher et al., 1993, 2001; Rivier and Clarke, 1997; Morosan et al., 2001), and functional imaging studies have shown that most complex sounds produce activation in PAC and surrounding areas in all normal listeners. The sensitivity of the cortical data means that we can investigate whether there are consistent differences between individual listeners in the location of functional activation within auditory cortex, and whether the differences correspond to differences in the sulcal and gyral morphol-

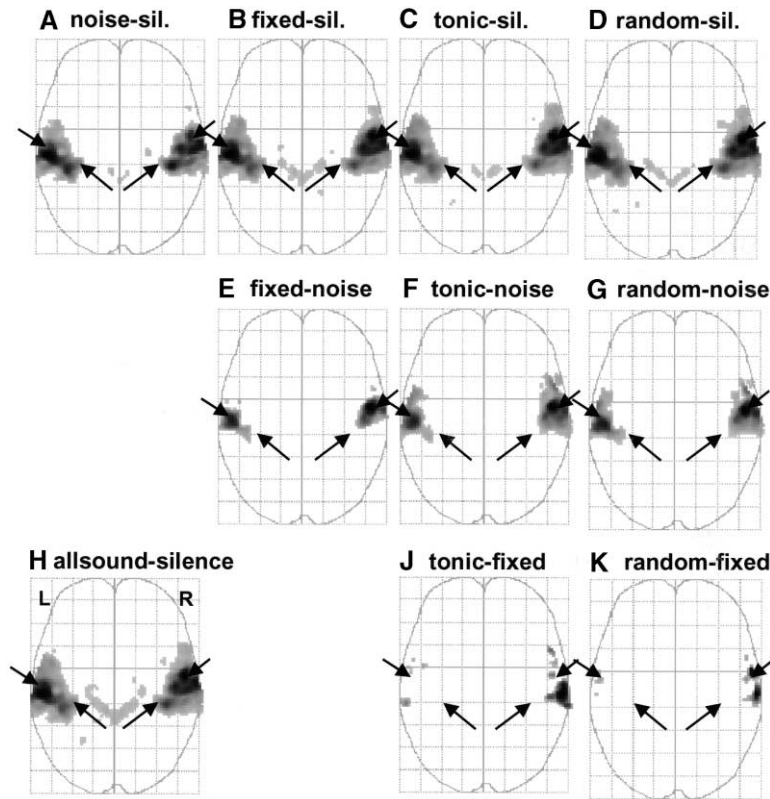


Figure 2. Axial Projection of Group Activation in Glass Brain View for Various Contrasts
Top row: (A) noise versus silence, (B) fixed pitch versus silence, (C) diatonic melody versus silence, and (D) random melody versus silence.
Middle row: (E) fixed pitch versus noise, (F) diatonic melody versus noise, and (G) random melody versus noise.
Bottom row: (H) all sound conditions contrasted with silence. (J) diatonic melody versus fixed pitch and (K) random melody versus fixed pitch. The height threshold for activation was $t = 5.00$ in every case ($p < 0.05$ corrected for multiple comparisons across the whole volume). The arrows mark the approximate position of Heschl's gyrus in the two hemispheres. The top row shows that these four types of sound produce activation in essentially the same cortical areas.

ogy of the individuals (Penhune et al., 1996; Leonard et al., 1998).

Results

The anteriormost, transverse temporal gyrus of Heschl was identified in each of our listeners, and there was good agreement between this specification of the location of HG in our listeners and that obtained in other studies. The details of the analysis are presented in Experimental Procedures. The group activation results are presented first with respect to the average position of HG for the group. Then the variability of the activation across listeners is compared to the variability of HG across listeners.

Regions of Activation in the Group

All Sounds versus Silence

The activation produced by all four sound conditions was compared to activation produced in the silence condition to illustrate the domain of cortical sensitivity to sound; the contrast includes 2592 volumes from all nine listeners (fixed-effects analysis). In cortex, this contrast yields bilateral activation in two large clusters shown in of Figure 2H (all sound-silence). The clusters are centered in the region of HG and PT, as would be expected. Outside this region, there is essentially no other cortical activation, perhaps because it was a passive listening experiment. The "V" of activity in the center of the panel is the subcortical activity reported in Griffiths et al. (2001). The position of HG for the group of listeners falls along the line between the arrowheads in

each hemisphere. The group activation is centered on the mean position of HG in the right hemisphere and along the posterolateral side of HG in the left hemisphere.

Individual Sound Conditions versus Silence

The individual sound conditions produced very similar patterns of activation when compared to silence, as shown in Figures 2A–2D. In the left hemisphere, there is little to distinguish among the four contrasts in terms of the region of activation; in the right hemisphere, the three sounds with pitch produce slightly more activity in the region just anterior to the lateral end of HG. Many of the peaks in these contrasts have t values above ten, ranging in some cases up to 40, and these peaks appear with remarkable consistency in all of the contrasts involving sound and silence. There are significant differences between conditions, but they are largely associated with different levels of activation at fixed positions within the main clusters rather than changes in the positions of peaks.

Differential Sensitivity to Pitch

To reveal regions associated with the processing of tonal sounds as opposed to noise, we examined the contrasts fixed versus noise, diatonic versus noise, and random versus noise (Figures 2E–2G). In the fixed versus noise comparison, the most prominent area of differential activity was in lateral HG in both hemispheres. The diatonic versus noise and random versus noise contrasts both exhibited very similar activation to fixed versus noise in HG. Anterior to HG, in the lateral part of PP, diatonic versus noise and random versus noise revealed bilateral activation that did not appear in the fixed versus

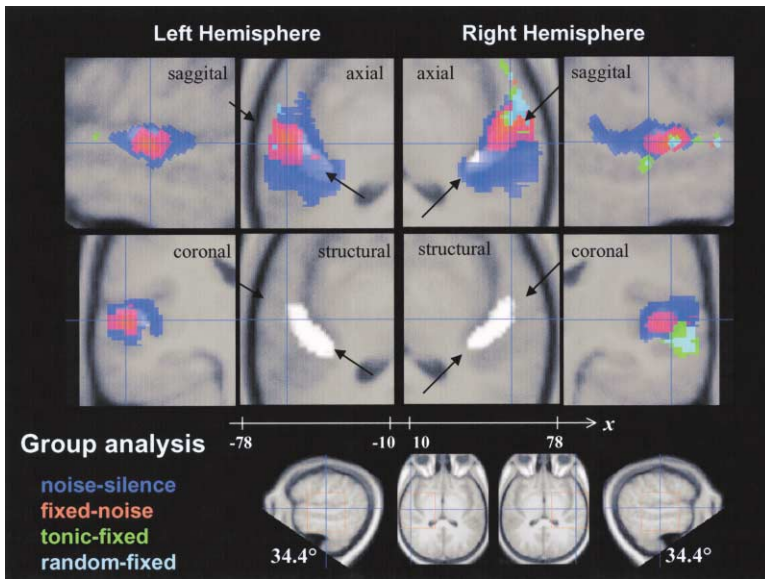


Figure 3. Group Activation for Four Contrasts Using a Fixed-Effects Model, Rendered onto the Average Structural Image of the Group

The height threshold for activation was $t = 5.00$ ($p < 0.05$ corrected). Blue, noise versus silence; red, fixed pitch versus noise; green, diatonic melody versus fixed pitch; cyan, random melody versus fixed pitch. The white highlight shows the position of Heschl's gyrus (HG) (the average of the HG maps for the nine individuals). The arrows show the midline of HG separately in each hemisphere. The position and orientation of the sections is illustrated in the bottom panels of the figure. The sagittal sections show front to the left for the left hemisphere and front to the right for the right hemisphere.

noise contrast. With regard to melody processing, these results suggest that there are a number of centers in auditory cortex (HG and PT) that process all stimuli coming up from subcortical structures in both hemispheres, and one of these regions in lateral HG is differentially active in the presence of pitch. We also inverted the three pitch versus noise contrasts to determine whether there might be a region particularly concerned with the processing of noise. No candidates were found; all of the regions activated by noise were activated to at least the same degree by sounds with pitch.

Differential Sensitivity to Melody

To reveal regions associated with melody processing, we examined the contrasts diatonic versus fixed and random versus fixed (Figures 2J and 2K). Both contrasts reveal differential activation to melody in STG and PP, but in this case, the activation is asymmetric with more activity in the right hemisphere. In lateral HG and medial HG, there is virtually no differential activation when compared with the previous contrasts, indicating that melody produced about the same level of activity as fixed pitch in HG. This suggests that HG is involved in short-term rather than longer-term pitch processing, such as determining the pitch value or pitch strength rather than evaluating pitch changes across a sequence of notes.

Diatonic versus Random Melody

In an attempt to identify regions that might be specifically involved in processing diatonic melodies, we examined the contrasts diatonic versus random and random versus diatonic. Neither contrast revealed significant peaks in any region of the brain. As a result, the data from the two melody conditions will be considered together in most of the following discussions.

Hierarchy of Melody Processing

A summary of the results to this point is presented in Figure 3. The structural and axial sections show the activity in a plane parallel to the surface of the temporal lobe and just below it; the sagittal sections are orthogonal to the axial sections and they face outwards so that in both cases, the view of the temporal lobe is from outside

the head. The highlighted regions in the structural sections show the average position of HG in the two hemispheres; these regions are replotted under the functional activation in the axial sections above the structural sections. The functional activation shows that as a sequence of noise bursts acquires the properties of melody (first pitch and then changing pitch), the region sensitive to the added complexity changes from a large area on HG and PT (blue) to a relatively focused area in the lateral half of HG (red) and then on out into surrounding regions of PP and STG (green and cyan mixed). The orderly progression is consistent with the hypothesis that the hierarchy of melody processing that begins in the brainstem continues in auditory cortex and subsequent regions of the temporal lobe. The activation is largely symmetric in auditory cortex and becomes asymmetric abruptly as it moves on to PP and STG with relatively more activity in the right hemisphere.

Variability in Anatomy and Functional Activation across Listeners

In this section, we examine how the anatomy of Heschl's gyrus in individuals relates to the individual's pattern of functional activation, and how the functional data of individuals relates to the pattern of activation observed in the group data.

The analysis of the anatomy of HG was summarized for each listener in terms of three points: (1) the centroid of the complete volume of HG, (2) the position of the medial end of HG, and (3) the position of the lateral end of HG (see Experimental Procedures). The group centroids for each of these points are presented in Table 1 with the standard deviations; the table shows that, following normalization, the variability in the position of HG is minimal. Specifically, the standard deviations for the medial and central centroids are less than one voxel (2 mm) in all three dimensions, and the standard deviations for the lateral centroid are less than two voxels on average.

With regard to the functional activation, the positions

Table 1. Coordinates of Anatomical Landmarks and Functional Activation Peaks in Heschl's Gyrus

	Left			Right		
	x	y	z	x	y	z
Medial end of HG	-35.8 (1.1)	-26.9 (1.8)	10.5 (1.0)	36.0 (1.8)	-26.3 (1.8)	10.3 (1.4)
Central HG	-49.1 (1.7)	-18.9 (2.0)	6.2 (1.0)	48.8 (1.3)	-16.8 (1.8)	5.4 (1.2)
Lateral end of HG	-62.1 (2.6)	-9.7 (3.7)	4.8 (1.4)	61.2 (3.7)	-7.8 (3.6)	3.3 (2.3)
Fixed versus noise (Individual listeners)	-55.3 (3.8)	-12.9 (4.6)	1.5 (2.9)	57.2 (2.5)	-8.8 (5.0)	-1.9 (3.3)
Penhune et al., 1996	-45.0	-19.8	7.8	47.8	-15.5	6.8
Morosan et al., 2001 (SPM normalized)	-44.6	-17.9	8.0	49.3	-12.9	7.2
Rademacher et al., 2001	-42.0	-21.0	7.0	46.0	-13.0	8.0

Coordinates of Heschl's gyrus. The top three rows show the mean position of the center of Heschl's gyrus and the medial and lateral ends of Heschl's gyrus, as determined from structural scans for the individual listeners. The fourth row shows the mean position of the peaks of activation for the contrast fixed versus noise from individual listeners. The bottom three rows show the location of Heschl's gyrus obtained from previous studies.

of the major peaks are very consistent across conditions within individuals. Nevertheless, the major peaks in the group data do not coincide with those in the data of individuals. Indeed, there are essentially no peaks in the group data that appear consistently for individuals. So, the position of a peak in the group data represents a location where activation from individuals overlaps in some way; it is not the location where a majority of the individuals all exhibit the same peak. In order to understand the form of the variability across listeners, axial and sagittal sections like those in Figure 3 were prepared for all nine listeners. The sections are presented with one listener per row in Figure 4. The first thing that the figure reveals is that the regions of activation in individuals are more focal than in the group average, indicating that the larger regions of activation in the group data represent overlapping focal regions from individuals—regions which in an individual are highly consistent across conditions. Moreover, variability in the location of activation differs in the three contrasts: noise versus silence (blue), fixed versus noise (red), and melody versus fixed (green). Accordingly, the variability is analyzed separately for each contrast.

Noise versus Silence

Noise produces foci of activation in the region of HG and/or PT in all listeners. The activation appears along the posterolateral edge of HG for some listeners (2, 3, 4, and 7), but not others (1, 5, 6, and 8). There is a region of activation at the posteromedial end of HG for some listeners (1, 2, 4, and 5), but not others (3, 6, and 8), and there is a concentration of activation anterior to HG in some listeners (3, 5, and 7), but not others. In short, the noise activation in individuals is restricted to focal regions of HG and PT, and together these regions produce the larger noise versus silence region in the group data. Within this larger region, however, the distribution of functional activation is quite variable across listeners. (The data of listener 9 are included in Figure 4 for completeness, but they are omitted from this analysis because the pattern of activation is so different from that of the other eight listeners.)

This pattern of variation in the noise activation stands in marked contrast to the consistency of peaks across

conditions within listeners; for any one of the small blue regions in an individual's data, it is typically the case that the region is activated by all of the different sounds used in the experiment, and the peak in the region is often in exactly the same place in each condition. Moreover, the degree of activation is the same, inasmuch as these regions rarely appear in contrasts involving one of the pitch-producing sounds and noise, nor do they appear when these contrasts are inverted. The obvious hypothesis is that these regions of activation represent centers that analyze broadband sounds for specific features or properties other than pitch and they are in somewhat different places in different listeners.

Fixed versus Noise

Within the region on lateral HG where sounds with pitch produce more activation than noise (red in Figure 2), individual listeners have, on average, four significant peaks in the left hemisphere and five in the right hemisphere. The peaks are shown in red in Figure 4 (Again, listener 9 is an exception. The contrast yielded no significant activation on the left and only one peak in a very anterior region on the right; accordingly, listener 9 was omitted from this analysis as well.). There is some variation in the pattern of activation; on the right, the activity for listeners 1, 6, and 7 appears to be slightly anterior to HG, and on the left, the activity for listeners 4, 5, 6, and 8 appears to be slightly posterior to HG. However, the region is small relative to the relatively large regions of activation produced by the noise versus silence contrast.

Centroids were calculated for the fixed versus noise peaks in each hemisphere for each listener and averaged across listeners; their coordinates are presented in the fourth row of Table 1. The functional centroid for the group is between the central and lateral anatomical centroids and a little below the line of anatomical centroids for the group. The standard deviations for the functional centroids are a little greater than those for the anatomical centroids at the lateral end of HG but, in general, the variability of the functional centroids is comparable to that of the anatomical centroids for the fixed versus noise contrast. This relatively small, bilateral region in lateral HG would appear to be a prime

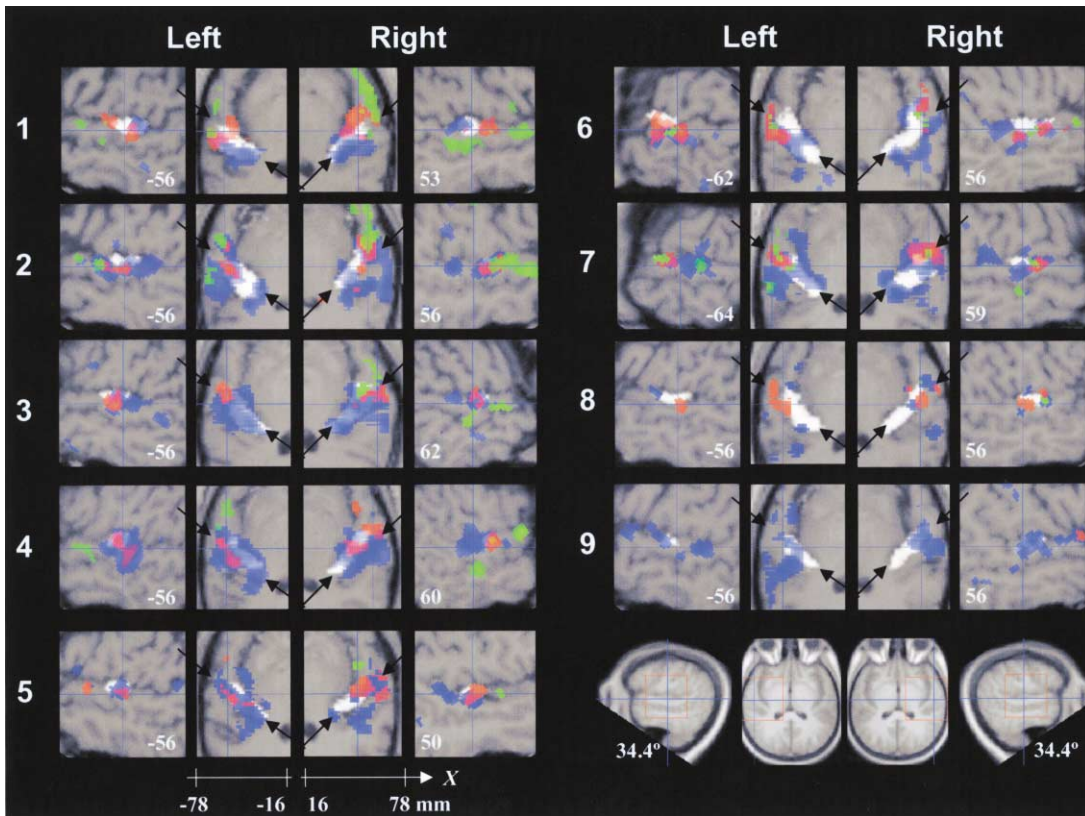


Figure 4. Activation for Three Contrasts in Individual Listeners (1–9) Rendered on Sections of Their Individual Structural Images

The orientation of the axial sections is the same as in Figure 3. The height threshold for activation was $t = 5.00$ ($p < 0.001$ uncorrected). Clusters with fewer than eight voxels were excluded from the figure to focus attention on the more significant clusters. The plane of each sagittal section is given in mm in Talairach-like space in each of the respective panels. The position of the individual's HG is highlighted in each case. The pairs of black arrows show the position of HG in the group. Blue, noise versus silence; red, fixed pitch versus noise; green, diatonic and random melodies versus fixed pitch.

candidate for a pitch center or tone processing center in auditory cortex.

Melody versus Fixed

There are three regions where melody produces more activation than fixed pitch: one at the lateral end of HG, one in STS, and one in PP. Figure 4 shows that there is somewhat more of this activity (green) in the individual data than in the group data (green and cyan in Figure 3), and the asymmetry is less pronounced in the individual data. There are two reasons for this: (1) the height threshold was reduced for the individual data ($p < 0.001$ uncorrected), and (2) the location of the activity is highly variable across listeners, so much of it does not appear in the average. With regard to asymmetry, for five of the nine listeners (1, 2, 3, 4, and 6), there is more activation in the right hemisphere than the left; the activation is roughly balanced for listener 7, and listeners 5, 8, and 9 exhibit little or no melody versus fixed activity. In the right hemisphere, six of the nine listeners have two regions of melody activation, but none has activity in all three of the regions identified in the average data, and no one combination of areas is overly common. In the left hemisphere, there is generally less activity and there is no discernible pattern. So, the distribution of activation associated with the processing of melody information outside auditory cortex is considerably more vari-

able across listeners than the processing of pitch information in lateral HG.

Melody produces essentially the same amount of activation as noise in PT and medial HG; in lateral HG, melody produces only slightly more activation than fixed pitch, and so the regions of differential activation to melody (green) are largely separate from those associated with noise and fixed pitch. The exception is in lateral HG, where listeners 2, 3, 6, and 7 exhibit some differential activation to melody within the region of activation associated with fixed pitch. It may be that lateral HG is involved in detecting when a pitch changes and the regions in PP and STS are involved in evaluating the pitch contour on a longer time scale.

Summary of Listener Variability

Eight of the nine listeners exhibit the hierarchy of melody processing observed in the group analysis. Within individuals, the locations of the peaks associated with the processing of noise and fixed pitch are highly consistent across stimulus conditions. Across listeners, however, the locations of the peaks associated with noise vary considerably. Similarly, the locations of peaks associated with the processing of melody in PP and STG vary considerably across individuals. The variability in the location of HG is small relative to the functional variability, and so the variability in the functional activation cannot be explained in terms of the gross anatomy.

Discussion

With regard to the hierarchy of melody processing: the fact that the contrast between fixed pitch and noise activates a region lateral to PAC on HG and the fact that the contrast between melody and fixed pitch produces activation outside of HG and PT supports the hypothesis that there is a hierarchy of pitch processing in human cerebral cortex with the activation moving anterolaterally as processing proceeds. There are limited anatomical data concerning connectivity in human cerebral cortex, but the relevant studies (Howard et al., 2000; Tardif and Clarke, 2001; Hackett et al., 2001) are consistent with the hierarchy of processing suggested by the functional data. Anatomical studies of the macaque show that auditory cortex is composed of a central core, a surrounding belt, and a lateral parabelt (Hackett et al., 1998) and that this cortical system is connected hierarchically (Kaas and Hackett, 2000). The core contains three regions: A1; a more rostral region, R; and an even more rostrot temporal region, RT (e.g., Kaas and Hackett, 2000). A recent comparative study of auditory cortex in macaques, chimpanzees, and humans (Hackett et al., 2001) suggests that core in macaques corresponds to the central three quarters of HG in humans. In our functional data, the centroid of the activity associated with pitch on HG (shown in red in Figure 3) is somewhat lateral and anterior to the region of HG traditionally associated with PAC. This suggests that the pitch region corresponds to the R or RT region of core. It should be noted, however, that the centroid of the pitch center in our study is below the central axis of HG, which may mean that it is not R or RT. The extra activation produced by melodies in STG and lateral PP is outside the core area; the region of activation in STG may correspond to a parabelt region in macaque, but the region in lateral PP seems rather anterior to be a parabelt region. In any event, the functional hierarchy of melody processing revealed in humans would appear to be consistent with the hierarchy of anatomical connections reported for macaques.

With regard to the emergence of asymmetry in the hierarchy: in the study of subcortical activation (Griffiths et al., 2001), the fixed versus noise contrast revealed symmetric activation in the brainstem and slightly asymmetric activation in the auditory thalamus with greater activation on the right; together these findings led to the conclusion that subcortical processing of temporal regularity was largely symmetric. With regard to auditory cortex, the PET study of Griffiths et al. (1998) revealed bilateral activation that increased with the strength of the pitch produced by the RI sounds, although the activation was somewhat asymmetric within respect to HG. It was centered on HG in the right hemisphere and toward the lateral end of HG in the left hemisphere. The data from the current study indicate that all the broadband stimuli produce bilateral and relatively symmetric activity in auditory cortex (HG and PT) (the blue and red regions in Figure 3). A specific test for hemispheric asymmetry was performed for the fixed versus noise contrast associated with pitch processing in lateral HG (as before, listener 9 was omitted from the analysis). The t value for the most prominent peak in this contrast was identified for each of the listeners in each hemisphere. The average values were 6.3 and 6.8 in the left

and right hemispheres, respectively, and the difference was not significant ($t = 1.47$, $df = 7$, and $p = 0.19$). For all eight listeners, prominent peaks occurred in the melody versus noise contrast and the random versus noise contrast at precisely the same positions as in the fixed versus noise contrast, and so the analysis was extended to these peaks as well, since they were assumed to represent the operation of the pitch extraction mechanism in the melody conditions. The average values for the melody versus noise contrast were 8.3 (left) and 9.3 (right), and the difference was not significant ($p = 0.39$). The average values for the random versus noise contrast were 8.2 (left) and 9.2 (right); the difference was not significant ($p = 0.09$). This suggests that pitch processing is largely symmetric in the hierarchy up to and including lateral HG.

Pronounced asymmetries are largely limited to the melody versus fixed contrast in the current study, and they emerge as the activation moves from HG out onto PP and STG. This interpretation is consistent with a host of neuropsychological studies. One of the earlier ones (Samson and Zatorre, 1988) shows that patients with anterior temporal-lobe resections (sparing HG) are more impaired when detecting a single note change in a pattern of three notes if the resections are on the right. A recent study (Johnsrude et al., 2000) indicates that patients with anterior temporal-lobe excisions that encroach upon the anterolateral extremity of HG in the right hemisphere, but not in the left, are more impaired when discriminating the direction of a melodic contour. Similarly, functional neuroimaging studies, which show relatively more activity in the right hemisphere in response to melodic sounds, support the hypothesis that the asymmetry is greater in regions anterior to auditory cortex (e.g., Zatorre et al., 1994; see Zatorre et al., 2002, for a review).

In their PET study, Griffiths et al. (1998) performed an interaction analysis to find areas where activity increased as a function of pitch strength, more for melodies than for fixed pitch, and found two pairs of relatively lateral regions—a posterior pair near the intersection of PT and STG (-58 , -42 , -2 and 72 , -40 , 6) and an anterior pair on PP (-54 , 10 , -18 and 58 , 12 , -26). The contrast between melody and fixed pitch in the current fMRI experiment is reasonably comparable, and Figures 2J and 2K show activation in similar regions to those in the PET study; the peaks in the posterior regions are at -62 , -28 , 2 and 66 , -30 , -2 and those in the anterior regions are at -56 , 6 , -10 and 54 , 14 , -16 . The activation in the fMRI study is more asymmetric than in the PET study, perhaps because it had greater spatial resolution.

There is also a recent study that relates musical ability, magnetoencephalographic (MEG) responses to modulated sinusoids, and gray matter volume in HG (Schneider et al., 2002). They found that the volume of gray matter in HG was significantly greater in a group of professional musicians than amateur and nonmusicians, and the professionals had stronger MEG responses in medial HG. The data were symmetric for amateur musicians and nonmusicians (consistent with our data), but there was a small, significant asymmetry for professional musicians. So for professional musicians, the asymmetry associated with melody may emerge one stage earlier in the hierarchy of melody processing.

The Hemispheric Specialization Hypothesis

In their review of asymmetry in response to speech and musical sounds, Zatorre et al. (2002) conclude that the processing of sounds with musical pitch results in relatively more activity in the right hemisphere, whereas the processing of sounds with critical timing information, like syllables with plosive consonants, results in relatively more activity in the left hemisphere. They then go on to propose that the auditory system has developed "... parallel and complementary systems—one in each hemisphere—specialized for rapid temporal processing (*left*) or for fine spectral processing (*right*) respectively" (page 40), and they draw an analogy with the uncertainty principle as it applies to time and frequency constraints in the spectrogram (their Box 2). Note that the discussion of the data is in terms of pitch, while the HS hypothesis and the analogy are described in terms of spectral processing. The activation associated with the pitch and melody of RI sounds appears to be largely compatible with the pitch and melody activation discussed by Zatorre et al. (2002), as noted above. It seems somewhat difficult, however, to reconcile the processing of RI sounds with the part of the HS hypothesis that says that the right hemisphere is specialized for fine spectral processing.

The HS hypothesis suggests that pitch is the result of fine spectral processing in auditory cortex or, to be more specific, that pitch arises from the detection of harmonically related peaks in the tonotopic representation of the Fourier magnitude spectrum of the sound as it occurs in or near auditory cortex (the concept is illustrated in Figures 1A and 1B of Griffiths et al., 1998). This cannot be the case for the RI sounds in this study, since there are no harmonically related peaks across the tonotopic dimension of the representation at any level in the auditory system (Figure 1G of the current paper). Moreover, the fine-grain timing information observed in physiological responses at the level of the brainstem is not observed in auditory cortex. This is what led us to suggest that the differential activation in lateral HG is associated with the calculation of precise values for pitch and pitch strength from the heights and widths of the peaks in a representation something like that shown in Figure 1K. It may be that the same area calculates precise spectral pitch values when the tonotopic representation in auditory cortex exhibits peaks and that the specialization is not so much one of fine spectral processing in auditory cortex to extract an accurate pitch estimate, but more one of accurate monitoring of the pitch information flowing from auditory cortex in subsequent centers concerned with whether the sound has the kind of stable pitch exhibited by musical notes, and what the intervals are when the pitch jumps from one note to another. A modified version of the HS hypothesis in which the specialization involves fine pitch tracking rather than fine spectral processing would appear to be in good agreement with the data, but in this case, the analogy with the uncertainty principle would seem somewhat tenuous.

Conclusions

All of the broadband sounds in this study produced activation bilaterally in a number of centers in HG and

PT independent of whether they produced a pitch or whether the pitch was changing over time. The fixed-pitch stimuli produced more activation than noise in lateral HG, bilaterally. The noise and fixed-pitch stimuli did not produce any substantial regions of activation outside HG and PT. When the pitch was varied to produce a melody, the sound produced additional, asymmetric activation in STG and lateral PP with relatively more activity in the right hemisphere.

While it is not possible at this time to be precise about the mechanics of auditory information processing at each stage, if we assume that there are three stages of melody information processing as proposed in the introduction, and they occur in the order specified, then the broad mapping from stage of processing to brain region would appear to be as follows. (1) The extraction of time-interval information from the neural firing pattern in the auditory nerve and the construction of time-interval histograms (e.g., the rows of Figures 1C and 1H), probably occurs in the brainstem and thalamus. (2) Determining the specific value of a pitch and its salience from the interval histograms probably occurs in lateral HG (e.g., by producing a summary histogram as in Figures 1E and 1K and locating the first peak). (3) Determining that the pitch changes in discrete steps and tracking the changes in a melody probably occurs beyond auditory cortex in STG and/or lateral PP. It would appear to be these latter processes associated with melody processing rather than pitch extraction per se that give rise to the asymmetries observed in neuropsychological and functional neuroimaging studies.

Experimental Procedures

Subjects

Nine normal-hearing listeners volunteered as subjects after giving informed consent (six male, three female, mean age 34.3 ± 8.9 years). None of the listeners had any history of hearing disorders or neurological disorders.

Stimulus Generation

The stimuli were sequences of noise bursts and regular-interval (RI) sounds. A RI sound is created by delaying a copy of a random noise and adding it back to the original. The perception has some of the hiss of the original random noise and also a weak pitch with a frequency at the inverse of the delay time. The strength of this pitch increases when the delay-and-add process is repeated (Yost et al., 1996). When the pitch is less than about 125 Hz and the stimuli are high-pass filtered at about 500 Hz, the RI sounds effectively excite all frequency channels in the same way as random noise (Patterson et al., 1996); compare the frequency profiles in Figures 1G and 1B. The perception of pitch in this case is based on extracting time intervals rather than spectral peaks from the neural pattern produced by the RI sound in the auditory nerve; compare the frequency profile in Figure 1G with the time-interval profile in Figure 1K. Quantitative models of the pitch of RI sounds based on peaks in the time-interval profile have proven highly successful (e.g., Pressnitzer et al., 2001).

There were five conditions in the experiment: four sound conditions and a silent baseline. The sounds were sequences of 32 notes played at the rate of four notes/s (8 s total duration). Each note was 200 ms in duration and there were 50 ms of silence between successive notes. The sounds were (1) random noise with no pitch (noise) and three RI sounds in which the pitch was (2) fixed for a given 32-note sequence (fixed), (3) varied to produce novel diatonic melodies (diatonic), or (4) varied to produce random note melodies (random). The pitch range for the diatonic and random melodies was 50 to 110 Hz. The pitch in the fixed-pitch sequences was varied

randomly between sequences to cover the same range as the melodies over the course of the experiment. All of the sounds were band-pass filtered between 500 and 4000 Hz using fourth-order Butterworth filters and presented to both ears at 75 dB SPL through magnet-compatible, high-fidelity electrostatic headphones (Palmer et al., 1998).

fMRI Protocol

Sparse temporal sampling was used to separate the scanner noise and the experimental sounds in time (Edmister et al., 1999; Hall et al., 1999). Blood oxygenation level-dependent (BOLD) contrast-image volumes were acquired every 12 s, using a 2.0-T MRI scanner (Siemens, VISION, Erlangen) with gradient-echo-planar imaging (TR/TE = 12,000 ms/35 ms). A total of 48 axial slices were acquired covering the whole brain. Each condition was repeated 48 times in random order. A T1-weighted MPRAGE high-resolution ($1 \times 1 \times 1.5$ mm) structural image was also collected from each subject on the same MR system. Further details are presented in Griffiths et al. (2001).

Data Processing and Analysis

Structural and functional data were processed and analyzed using SPM99 (<http://www.fil.ion.ucl.ac.uk/spm>). The BOLD time series was realigned to the first image of the series and then the structural image was coregistered to these images and resampled to $2 \times 2 \times 2$ mm resolution. The realigned BOLD images were normalized to the standard SPM EPI template (ICBM 152, Brett et al., 2002) using affine and smoothly nonlinear spatial transformations: the same normalization parameters were applied to the structural images. The resulting images are in standardized, Talairach-like space. The ICBM-152 template creates images that are a few millimeters displaced from the Talairach brain (Talairach and Tournoux, 1988), particularly in the superior-inferior (z) dimension (Brett et al., 2002). This is the reason for the phrase Talairach-like. Finally, the functional data were smoothed with a Gaussian filter of 5 mm (full-width at half maximum). Fixed-effects analyses were conducted on each listener's data (288 scans) and across the whole group of subjects (with a total of 2592 scans) using the general linear model. The height threshold for group activation was $t = 5.00$ ($p < 0.05$ corrected for multiple comparisons across the whole volume). It was reduced somewhat for the individual data to $t = 5.00$ ($p < 0.001$ uncorrected).

Evaluation of Individual Anatomy

In order to observe the relationship between functional activation and macroscopic anatomy, the putative location of primary auditory cortex was identified in each listener. All four of the authors labeled the first transverse temporal gyrus (HG) in both hemispheres of each listener, using a software labeling package (MRIcro: <http://www.psychology.nottingham.ac.uk/staff/crl/micro.html>) and the anatomical criteria suggested by Penhune et al. (1996). Both white and gray matter were included in the labeled volumes. In some listeners HG is duplicated or partially duplicated (Penhune et al., 1996; Leonard et al., 1998). In this study, the duplication was ignored, and the labeling was restricted to the part of HG anterior to any dividing sulcus (Rademacher et al., 1993; 2001). So our estimates of the extent of primary auditory cortex are a little more conservative than is traditional (Rademacher et al., 2001).

The labeling of the four judges was combined to produce a labeled volume of HG for each listener by including all voxels that two or more judges had labeled as part of HG. The result is a three-dimensional HG map that is coregistered with the individual's functional data. Finally, a mean HG volume was created for the group by averaging across the nine individually labeled volumes.

Comparison of Anatomy with Previous Studies

The location of HG obtained in the current study was compared with that obtained in one morphological study (Penhune et al., 1996) and one cytoarchitectonic study (Morosan et al., 2001; Rademacher et al., 2001): in the former, HG was identified on the MR scans of 20 individuals; in the latter, PAC was identified in brains of ten individuals. Probability volumes of HG were obtained from Penhune, and centroids were calculated for the volumes in a probability-weighted fashion (Hall, et al., 2002). Rademacher et al. (2001) report

centroids for a normalized version of their PAC data (p. 677). To ensure comparability, we obtained the raw data for nine cases from Rademacher and colleagues, and we applied the default, smoothly nonlinear, normalization of SPM99 to the data. Centroids were then calculated for these maps using the same probability-weighted function applied to the Penhune maps. All three sets of centroids are listed in Table 1. Note that the estimates based on HG and PAC data are almost identical.

Overall, the positions for HG estimated from the current study are in good agreement with the positions derived in previous studies. The group centroid in the current study is more anterior on the right than on the left (t (8df) = 3.69, $p < 0.01$), consistent with previous reports. A comparison with other studies shows a displacement of 5–7.7 mm in the left and 3.2–5.9 mm in the right hemisphere. There was a significant difference ($t = 2.71$, $p < 0.05$) in left/right displacement when compared with the centroids of HG, while this displacement was not significantly different from the estimates of the center of PAC (see Table 1). One-group t tests, testing for significant displacement (2-tailed) between left and right hemispheres separately for x , y , and z revealed that only the x coordinate in the left and the z coordinate in the right were significantly different in location from all three other estimates. Our x coordinate in the left is, on average, 5.2 mm (individual averages range 4.1–7.1 mm) more lateral than in the other estimates. Our z coordinate in the right is 2.0 mm inferior to the other estimates (individual average range 1.4–2.6 mm).

Acknowledgments

The research was supported by the UK Medical Research Council (G9900369, G9901257).

Received: May 23, 2002

Revised: September 24, 2002

References

- Brett, M., Johnsrude, I.S., and Owen, A.M. (2002). The problem of functional localization in the human brain. *Nat. Rev. Neurosci.* 3, 243–249.
- Edmister, W.B., Talavage, T.M., Ledden, P.J., and Weisskoff, R.M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Hum. Brain Mapp.* 7, 89–97.
- Ehret, G., and Romand, R. (1997). *The Central Auditory System* (New York: Oxford University Press).
- Griffiths, T.D., Büchel, C., Frackowiak, R.S.J., and Patterson, R.D. (1998). Analysis of temporal structure in sound by the brain. *Nat. Neurosci.* 1, 422–427.
- Griffiths, T.D., Uppenkamp, S., Johnsrude, I., Josephs, O., and Patterson, R.D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nat. Neurosci.* 4, 633–637.
- Guimares, A.R., Melcher, J.R., Talavage, T.M., Baker, J.R., Ledden, P., Rosen, B.R., Kiang, N.Y.S., Fullerton, B.C., and Weisskoff, R.M. (1998). Imaging subcortical activity in humans. *Hum. Brain. Mapp.* 6, 33–41.
- Hackett, T.A., Stepniewska, I., and Kaas, J.H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* 394, 475–495.
- Hackett, T.A., Preuss, T.M., and Kaas, J.H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.* 441, 197–222.
- Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., and Bowtell, R.W. (1999). "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Mapp.* 7, 213–223.
- Hall, D.A., Johnsrude, I.J., Haggard, M.P., Palmer, A.R., Akeroyd, M.A., and Summerfield, A.Q. (2002). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 12, 140–149.
- Howard, M.A., Volkov, I.O., Mirsky, R., Garell, P.C., Noh, M.D., Graner, M., Damasio, H., Steinschneider, M., Reale, R.A., Hind, J.E., and Brugge, J.F. (2000). Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol.* 416, 79–92.

- Johnsrude, I.S., Penhune, V.B., and Zatorre, R.J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155–163.
- Kaas, J.H., and Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. USA* 97, 11793–11799.
- Krumbholz, K., Patterson, R.D., and Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *J. Acoust. Soc. Am.* 108, 1170–1180.
- Leonard, C.M., Puranik, C., Kuldau, J.M., and Lombardino, L.J. (1998). Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? *Cereb. Cortex* 8, 397–406.
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., and Zilles, K. (2001). Human primary auditory cortex: subdivisions and mapping into a spatial reference system. *Neuroimage* 13, 684–701.
- Palmer, A.R., and Winter, I.M. (1992). Cochlear nerve and cochlear nucleus responses to the fundamental frequency of voiced speech sounds and harmonic complex tones. In *Auditory Physiology and Perception*. Y. Cazals, L. Demany, and K. Horner, eds. (Oxford: Pergamon), pp. 231–239.
- Palmer, A.R., Bullock, D.C., and Chambers, J.D. (1998). A high-output, high-quality sound system for use in auditory fMRI. *Neuroimage* 7, S359.
- Patterson, R.D. (1994). The sound of a sinusoid: time-interval models. *J. Acoust. Soc. Am.* 96, 1419–1428.
- Patterson, R.D., Allerhand, M., and Giguere, C. (1995). Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform. *J. Acoust. Soc. Am.* 98, 1890–1894.
- Patterson, R.D., Handel, S., Yost, W.A., and Datta, A.J. (1996). The relative strength of the tone and noise components in iterated rippled noise. *J. Acoust. Soc. Am.* 100, 3286–3294.
- Penhune, V.B., Zatorre, R.J., MacDonald, J.D., and Evans, A.C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb. Cortex* 6, 661–672.
- Pressnitzer, D., Patterson, R.D., and Krumbholz, K. (2001). The lower limit of melodic pitch. *J. Acoust. Soc. Am.* 109, 2074–2084.
- Rademacher, J., Caviness, V.S., Steinmetz, H., and Galaburda, A.M. (1993). Topographical variation of the human primary cortices and its relevance to brain mapping and neuroimaging studies. *Cereb. Cortex* 3, 313–329.
- Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.J., and Zilles, K. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* 13, 669–683.
- Rauschecker, J.P., and Tian, B. (2000). Mechanisms and streams for processing 'what' and 'where' in auditory cortex. *Proc. Natl. Acad. Sci. USA* 97, 11800–11806.
- Rivier, F., and Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *Neuroimage* 6, 288–304.
- Samson, S., and Zatorre, R.J. (1988). Melodic and harmonic discrimination following unilateral cerebral excision. *Brain Cogn.* 7, 348–360.
- Schneider, P., Scherg, M., Günter Dosch, H., Specht, H.J., Gultshalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694.
- Talairach, P., and Tournoux, J. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. (Stuttgart: Thieme).
- Tardif, E., and Clarke, S. (2001). Intrinsic connectivity in human auditory areas: tracing study with Dil. *Eur. J. Neurosci.* 13, 1045–1050.
- Wessinger, M., Van Meter, J., Tian, B., Pekar, J., and Rauschecker, J.P. (2001). Hierarchical organisation of the human auditory cortex revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* 13, 1–7.
- Yost, W.A. (1996). Pitch of iterated rippled noise. *J. Acoust. Soc. Am.* 100, 511–518.
- Yost, W.A. (1998). Auditory processing of sounds with temporal regularity: Auditory processing of regular-interval stimuli. In *Psychophysical and Physiological Advances in Hearing*, A. Palmer, A. Rees, Q. Summerfield, and R. Meddis, eds. (London: Whurr), pp. 546–553.
- Yost, W.A., Patterson, R.D., and Sheft, S. (1996). A time-domain description for the pitch strength of iterated rippled noise. *J. Acoust. Soc. Am.* 99, 1066–1078.
- Zatorre, R.J., Evans, A.C., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908–1919.
- Zatorre, R.J., Belin, P., and Penhune, V.B. (2002). Structure and functions of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46.